ARL-TR-7298 • MAY 2015

US Army Research Laboratory

Agent Transparency for an Autonomous Squad Member

by Michael W Boyce, Jessie YC Chen, Anthony R Selkowitz, and Shan G Lakhmani

NOTICES

Disclaimers

ARL-TR-7298 ● MAY 2015

US Army Research Laboratory

Agent Transparency for an Autonomous Squad Member

by Michael W Boyce
Oak Ridge Associated Universities (ORAU)

Jessie YC Chen
Human Research and Engineering Directorate, ARL

Anthony R Selkowitz and Shan G Lakhmani
University of Central Florida, Institute for Simulation and Training

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
May 2015	Final	1 February 2014–28 February 2015

4. TITLE AND SUBTITLE
Agent Transparency for an Autonomous Squad Member

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael W Boyce, Jessie Y C Chen, Anthony R Selkowitz, and Shan G Lakhmani

5d. PROJECT NUMBER
ARPI

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-HRM-AR
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7298

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The ability to understand the reasoning behind an intelligent agent's actions can help to increase operator performance as the use of human-agent teams for military operations grows. This experiment tested the effect of display design to convey environment and intelligent agent information in a simulation-based unmanned ground vehicle monitoring task. Three groups were tested with visual displays representing 1 of 3 types of information: current status only (group 1); current status with reasoning information (group 2); and current status, reasoning information, and projected information (group 3). Performance measures included comprehension of situation awareness probes, operator trust based on 3 different surveys, workload, and system usability. Results indicated a significant interaction between conditions and pre- and postadministration of a trust survey modified from Jian et al. (2000), with only group 2 increasing in trust preexperiment in comparison with postexperiment. The situation awareness probes failed to yield any significant differences among the conditions. No significant effects of operator workload or individual difference factors were observed across conditions. This research demonstrates the potential of agent transparency displays to improve Soldier trust and situation awareness.

15. SUBJECT TERMS
agent transparency, human agent teaming, display design, visualization, situation awareness

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Michael W Boyce
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	SAR	84	407-384-3910

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

This research was supported by the US Department of Defense's Autonomy Research Pilot Initiative. The authors wish to thank MaryAnne Fields, Daniel Barber, Erica Valiente, Jonathan Harris, and Susan Hill for their contribution to this project.

1. Introduction

1.1 Agent Transparency for an Autonomous Squad Member

A Soldier is accompanying his squad on a routine reconnaissance mission in a wooded, partially concealed area. A small display mounted to the Soldier's body armor begins to flash. This message is not coming from the commander or any other human in the environment. Instead, it is coming from an unmanned ground vehicle (UGV) that is moving with the team, which the Soldiers have brought with them to enhance their understanding of their surroundings. The Soldier looks down at his display and notices that the path ahead has been recently attacked by mortar fire. However, there is an alternative path that is protected due to the ledge of a rock formation. Knowing that there are potential troops nearby, the Soldier motions to the squad to take the alternate path, and the squad safely completes their mission.

Although this may seem like something out of a recent science fiction movie, the use of human-robot teams continues to grow in the military (Barnes and Evans 2010; Ososky et al. 2014). The environment in which dismounted Soldiers—those Soldiers not using a vehicle—is characterized by situations that require the Soldier to act quickly and effectively (Oron-Gilad et al. 2011). The advancement of robotic capabilities provides these Soldiers with the opportunity to assign specific job functions to the robot, while reserving others for the Soldier (Chen and Barnes 2014; Miller 2014). A collaboration is formed between the human and robot (Ososky et al. 2014); the robot is also referred to as an intelligent agent. An intelligent agent is a system that can observe and adjust actions based on the needs to achieve mission goals (Russell and Norvig 2009; Chen and Barnes 2014).

This experiment investigates agent transparency as applied to UGV displays. Agent transparency describes a display in which the agent's status, reasoning, abilities, and plans for future actions help comprehension by dismounted Soldiers (Chen et al. 2014). A major component of transparency is the shared intent and shared awareness between the 2 parties, according to the definition proposed by Lyons (Lyons 2013; Lyons and Havig 2014). The Soldier needs to be receiving feedback or information on how their actions are affecting the system's understanding of situation awareness (SA). What the Soldier needs is an adequate understanding of the complexity of the environment around them; this is also known as (SA), the topic of the following section.

1.2 Situation Awareness

According to Endsley (2012), at a very basic level humans need to understand what is going on in the situation around them. Formally defined, SA is "the perception of elements in the environment . . . the comprehension of their meaning, and the projection of their status in the near future" (Endsley 1988).

SA consists of 3 levels:

- Level 1 SA: the direct perceptual properties of the elements in the environment.
- Level 2 SA: merging those elements into a comprehensible picture.
- Level 3 SA: the upcoming states given the state of current elements, the reasoning behind those, and how it changes over time in relation to the mission goal (Endsley 2012).

In the current experiment, different visualizations would contribute to different levels of SA due to the way SA information is processed. SA encompasses both top-down and bottom-up processing. Top-down processing utilizes mental models of the world to classify appropriate actions to achieve the goal. Mental models, according to Rouse and Morris (1985), are frameworks and relationships developed in the mind to help understanding. Bottom-up processing focuses around the basic symbology of elements in the environment. Effective SA requires an active switching between bottom-up and top-down processing. By focusing only on the goal, an individual might not recognize something that has changed in the environment. By focusing solely on the elements in the environment, an individual might demonstrate attentional tunneling thereby losing sight of the overall goal (Endsley 2012; Endsley and Jones 2012). When an individual has good SA, they can better comprehend why an agent is behaving a certain way. Through understanding actions, the human can develop trust in the agent, which is discussed in the next section.

1.3 Trust

When humans are working collaboratively with an intelligent agent, one factor that contributes to performance is the level of trust between the person and the agent. Lee and See (2004) *define* trust as the user's "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability".

Applied to this experiment, information coming from an intelligent agent assists in decision making. If that information helps achieve the task more efficiently, the Soldier will continue to use and develop trust in that agent. Without trust, Soldiers may look at the agent as an increase in workload without an increase in performance.

Lee and Moray (1992) identified performance, process, and purpose as the 3 fundamental bases of human-automation trust or in this case, agent. Performance defines the current state and characteristics of the agent. Process describes how the agent achieves its necessary goals. Purpose refers to the human intent of what an intelligent agent was created to achieve. The degree to which these bases are effectively conveyed can affect levels of operator's trust.

There are other factors that change an operator's trust as well. Hancock et al. (2011) found that performance factors such as reliability and predictability were the strongest factors indicating trust (Hoffman et al. 2013). Lee and See (2004) developed a series of recommendations for trustable automation:

1. Appropriate trust is more important than higher levels of trust
2. Display past performance
3. Show the entire process of the automation including intermediate steps
4. Simplify the automation to make it easier to learn
5. Demonstrate the purpose in the context of the current goals of the operator
6. Educate operators on the reliability constraints and appropriate use of the automation.

These guidelines provide a foundation toward developing transparent interfaces, as well as trustable ones. They can be applied to all levels of SA, especially Level 2 (Comprehension), and Level 3 (Projection). It is from these principles and levels that the SA-Based Transparency model was created, which is discussed next.

1.4 SA-Based Agent Transparency Model (SAT Model)

The SAT Model (Chen et al. 2014), is a conceptual way of thinking for organizing transparency requirements related to an intelligent agent. The model consists of existing frameworks that supports understanding in dynamic environments by leveraging Endsley's (1988) model of SA (Perception, Comprehension, and Projection) as its foundation. The model integrates the belief, desire, and intention framework that is designed to support the architecture of intelligent agents (Rao and Georgeff 1995).

These transparency requirements exist in the SAT Model at 3 different levels. Level 1 consists of the current state, goals in the domain, and any existing action plans available to the agent. Level 2 explains the underlying reasoning that the agent uses to choose one decision over another. This decision takes place based on the context of the affordances and constraints in the environment. Level 3 provides information on the future state of the agent, as well as any uncertainty about what may occur to help educate the operator about potential impacts of available decision options. Figure 1 has a breakdown of activities according to SAT Model level.

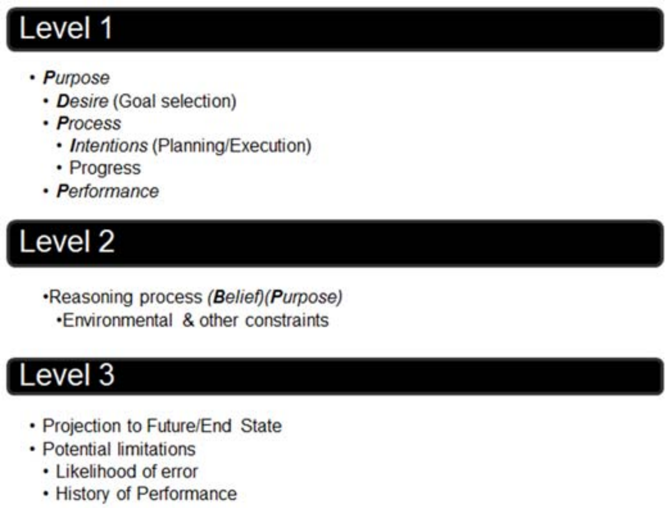


Fig. 1 SA-based agent transparency (SAT) model

Although the SAT Model is effective for organizing thoughts and ensuring information requirements are met, transitioning between theory and actual design is not an easy task, but creating these designs was critical to the current study. To develop the designs, the information processing model of Rasmussen (Rasmussen and Vicente 1989; Bennett and Flach 2011) was used, also known as the symbol, rule, and knowledge (SRK) framework, which is discussed further in the next section.

1.5 The SRK Framework

Rasmussen proposed 3 types of processing: skill based, rule based, and knowledge based (also known as signal, sign, and symbol representations) (Bennett and Flach 2011).

For skill-based or signal processing, an individual can directly interpret the environment. For rule-based or sign-based processing, human graphical interpretation relies on cultural or design conventions that are outside of direct perception. For knowledge-based or symbol processing the connection between the

symbol and its meaning requires interpretation. The relationship is ambiguous, and techniques like pattern recognition or distinguishing consistency are needed to differentiate between relationships (Bennett and Flach 2011). Rasmussen's work led to the development of ecological interface design (EID). They explain EID as "trying to make the interface transparent, that is, to support direct perception directly at the level of the user's discretionary choice. . ." (Rasmussen and Vicente 1990).

1.6 Current Study

This experiment simulated an intelligent agent monitoring environment. A dismounted Soldier had to understand the status of the autonomous squad member (ASM), a UGV. The role of the Soldier was to provide updates and information to the rest of the squad regarding the ASM's activities. The simulated vehicle was part of a scenario-based visual display. The participant had to answer questions about their understanding of the agent's activities based on environmental affordances and constraints. The amount of information displayed corresponded to the levels of the SAT model. The number of scenarios, questions, and waypoints was held constant throughout the experiment.

1.7 Stated Hypotheses/Objectives

This experiment manipulated the amount of transparency information of the ASM display to assist the monitor with comprehension (through SA probes) of the ASM's activities. There were 3 levels of transparency information:

1. Group 1: current status information
2. Group 2: adds environmental affordance/constraint regions
3. Group 3: adds visualization of projected status and uncertainty

Manipulation of displayed transparency information is presumed to influence operator's ability to maintain SA. Therefore, as transparency information increases, so too should operator SA increase.

> *Hypothesis (H) 1: Operator SA, as demonstrated through performance on the SA probes, will increase with the addition of each level of agent transparency information.*

Trust in an automated system can influence operators' perception of the situation. Three scales were also used to assess monitor trust. Increased agent transparency should positively influence operator trust in the automated system.

H2: Increased agent transparency will raise operator trust, as determined by change in trust or differences in posttask trust.

Increased transparency information requires more effort on the part of the operator to process. Consequently, increased transparency should influence operator workload.

H3: Workload, as measured through the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) will differ between agent transparency information conditions with more transparency information increasing workload.

The experiment tests effects of individual differences in mental rotation, Operational Span (OSPAN), attentional control, and prior gaming experience on the monitors' comprehension.

H4: Individual difference factors (mental rotation, gaming experience, OSPAN, and attentional control) will be significant covariates to percentage correct as measured by the SA probes.

Finally, this increased transparency is expected to influence operator's subjective usability of the automated system's interface.

H5: System usability, as measured through the system usability scale will increase with additional agent transparency information.

2. Method

2.1 Participants

Forty-eight participants signed up for the study through an online research signup system (Sona Systems). Exclusion criteria within Sona Systems ensured that all participants were college students, above the age of 18, and US citizens. All participants had to pass a color blindness test prior to beginning the experiment. No participants were found to be colorblind. The data of 3 participants from the study were not a part of the analysis. The 3 participants' data were incomplete, and therefore were used as pilot data. Therefore, the final number of participants was 45 (M_{age} = 21.04, SD_{age} = 2.17, 27 men, 17 women, 1 nondisclosed).

The participants were representative of several different areas of study: 12 were from Engineering, 9 from Business, 7 from Arts and Humanities, 6 from Biological Sciences, 5 from Social Sciences, 4 from Computer Science and 1 from Physical Sciences and Criminal Justice. Out of the 45 participants, 34 had less than 4 years of college, 10 had 4 years of college, and 1 had an advanced degree. The

participants reported an average 7.31 h of sleep the night prior to the experiment. Only one participant characterized himself as a novice computer user, with 10 of the participants reporting computer programming experience, and the rest proficient with at least one software package.

The setting of the research was in a dedicated experiment room with a divided environment between the participant and the research team member, and participants were provided with a computer system in front of them. A nonrecording camera was set up to allow the research team member to monitor the participant in the event they started to fall asleep. Participants were compensated \$15/h for their time, rounded up to the nearest half hour with each participant receiving a minimum of 1 h's pay, even if they did not complete the experiment. The study received approval from the Institutional Review Board of the US Army Research Laboratory.

2.2 Apparatus

2.2.1 Simulator

The scenario-based simulation task was to monitor a visual display that provided information on the actions of the ASM. The ASM was represented by a small vehicle icon that moved along a predefined path (Fig. 2). The surrounding environment contained areas that were hazardous as well as areas that would afford better ASM performance.

Fig. 2 Example of ASM display

2.2.2 Surveys and Tests

2.2.2.1 Demographics

A demographics questionnaire (Appendix A) was administered prior to the beginning of the training. Information included age, gender, education level, how familiar they were with technology and how often they reported playing video games. Video game frequency is represented by 6 groups; daily, weekly, monthly, every few months, rarely, and never. If the participants responded that they played either daily or weekly, they were categorized as frequent gamers. Participants were also categorized based on the types of games they played, as either action game players or action game nonplayers. Action games were defined as games with a time constraint, where the majority of challenges are physical tests of skill, requiring good hand-eye coordination and quick response times (Adams 2013).

2.2.2.2 Color Vision Test

An Ishihara color vision test (using 9 test plates) via PowerPoint presentation was a part of pre-experiment activities. The Ishihara color vision test was used because it was necessary to verify that individuals were not color-blind.

2.2.2.3 Mental Rotation

Mental rotation was assessed using the Vandenberg and Kuse Mental Rotation test (1978). The test, included as Appendix B, contains 24 items. Each item has a target figure followed by 2 reproductions of the target and 2 distractors. The participant has to select which 2 of the 4 figures are rotated representations of the desired target. Mental rotation was assessed because it has been shown to be a predictor of spatial ability when examining navigation-based tasks (Rehfeld 2006). Research has shown that mental rotation is 1of 4 cognitive operations required during navigation (Aretz and Wickens 1992).

2.2.2.4 Working Memory

Since this research requires the participants to remember information and then answer questions on SA, the OSPAN test (Engle 2002) was part of the pre-experiment testing. The OSPAN test assesses working memory capacity for both mathematical equations and a series of letters that participants are asked to remember.

2.2.2.5 Perceived Attentional Control

The participants' Perceived Attentional Control (PAC) was evaluated using the Attentional Control Survey shown in Appendix C (Derryberry and Reed 2002). PAC is an individual difference factor that can have an impact on attention focus and the ability to shift between tasks. The scale has been shown to have good internal reliability ($\alpha = 0.88$).

2.2.2.6 System Usability

The participant's perceived satisfaction with the user interface is measured using the System Usability Scale presented as Appendix D (Brooke 1996). The scale consists of 10 items with 5 response options ranging from strongly agree to strongly disagree, with scores ranging from 0 to 100. The scale has been shown to have good internal reliability ($\alpha > 0.90$). Perceived system usability was measured to determine that any differences between conditions were attributable to experimental manipulation rather than dissatisfaction with the interface.

2.2.2.7 Modified Jian Trust Scales

Participants were given 2 modified scales of the Trust in Automated Systems scale. Multiple versions were administered because although the content was similar, they were presented with alterations of the original scale. The first, included as Appendix E (Jian et al. 2000), was an 11-item scale administered both prior and following the observation of the autonomous agent to establish change throughout the experiment in their trust of the display. The scale consisted of semantic differential scales that rated from 1 to 7 (1 = Not at all, 7 = Extremely).

The second modified scale, shown in Appendix F and administered postexperiment, assesses trust of the system as it corresponds with the 4 stages of human information processing (Parasuraman et al. 2000). The 4 stages include information acquisition, information analysis, decision and action selection, and action implementation. These stages were conceptualized in the scale as gathering or filtering information, integrating and displaying analyzed information, suggesting or making decisions, and executing actions. The modified scale included 16 questions, each scored on a 1–7 (1 = Not at All, 4 = Neutral, and 7 = Extremely) Likert scale.

2.2.2.8 Schaefer Human-Robot Trust Scale

Participants were also given a shortened version of Schaefer's (2013) scale Appendixes G and H) on human-robot trust in a pre- and postformat. The scale consists of a 14-question rating scale ranging from 0 to 100 based on the percentage of time the robot will act in the desired manner. The participant takes the prescale

after viewing a picture of the robot. The prescale is meant to assess the predisposition for trust of the participant. The experimenter re-administers the scale postexperiment to assess the change in robot trust due to experimental manipulation.

2.3 Procedure

After being briefed on the purpose of the study, the participants signed the informed consent. Participants completed an Ishihara color vision test (with 9 test plates) via PowerPoint slides. They then completed a demographics questionnaire, an attentional control survey, a mental rotation survey, and working memory test. The experimental task consisted of monitoring the ASM through a simulated environment and answering SA probes throughout the course of the experiment. The participant is told he/she must monitor an ASM moving with a group of dismounted Soldiers. The participant has a start and end goal and needs to monitor scenarios consisting of 10 waypoints.

The participants were randomly assigned to 1 of the 3 experimental conditions (15 subjects per condition): group 1, group 2, and group 3. In the first condition (group 1), participants were provided with a current status icon representing 4 different resources of the ASM: perception, battery, mechanical, and communication (Fig. 3).

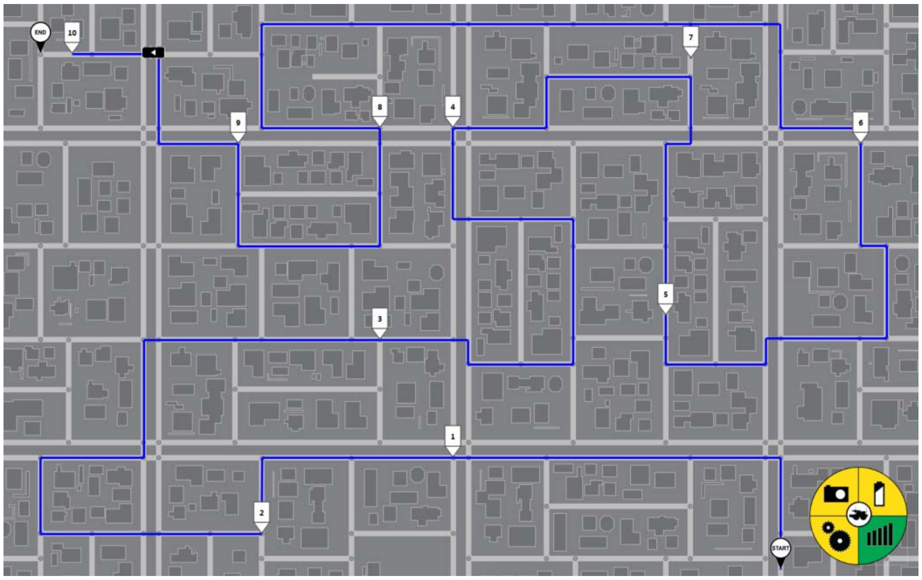


Fig. 3 ASM group 1 display

Each icon changed color (red, yellow, and green) as the mission progressed. Green meant that the resource was in good condition or full strength. Yellow meant that the resource was in average condition or moderate strength. Red meant that the resource was in poor condition or low strength.

In the second condition (group 2), environmental information was added for 3 environmental characteristics (shelling, communications, and terrain). Each characteristic was represented by a region, which displayed either an affordance or a hazard (Fig. 4).The triangles and circles that surround the regions do not have any specific meanings. What does matter is whether the color is red or green. A red shelling zone means the potential for enemy shelling, while a green shelling zone means the potential for fire support from friendly/allied units. A red communication zone means there are communication jamming devices in the area, and green communication zones mean areas of consistent and clear communication. A red terrain zone indicates the possibility of difficult or unpassable terrain, while a green terrain zone means an area of smooth or easy to traverse terrain.

Fig. 4 ASM group 2 display

In the third condition (group 3), uncertainty and projection information were added (Fig. 5). All of the environmental meanings from group 2 still hold here. The addition is the presence or absence of uncertainty, the environmental characteristics could be either certain or uncertain (represented by opacity level). For projection, a second icon set was added to represent projected resource amounts of the ASM. In the current status, an icon represents the present amount of a particular resource; in the projection status, an icon represents the expected end state when finished navigating through the scenario.

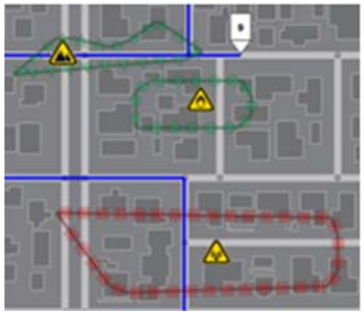


Fig. 5 Example displays for ASM projection and uncertainty information (See Fig. 2 for full display)

The task was to monitor a display that provided visual information on the actions of the ASM, as well as the potential hazards and affordances in the surrounding environment. Participants were given a brief training session to familiarize them with the display. Participants finished the training with a practice scenario that looked very similar to the experimental scenario with the only difference being placement of icons.

The training lasted approximately 45 min. In the experimental environment, the participants progressed through a series of 6 scenarios in the same order. In the scenarios, participants reported their comprehension of agent activities via SA probes. During each scenario, participants were prompted for their Instantaneous Workload Assessment (Jordan and Brennan 1992). The instantaneous workload assessment takes subjective workload assessments during the middle of a task. Following each scenario, the participants took the NASA-TLX (Hart and Staveland 1988). Participants also took 2 modified trust scales (one scale occurred pre- and postexperiment; the other only occurred postexperiment) based on the Trust of Automated Systems scale (Jian et al. 2000) as well as a shortened version of an existing trust scale on human-robot interaction postexperiment (Schaefer 2013). Participants finished by taking the System Usability Scale (SUS) (Brooke 1996), and the Vandenberg and Kuse Mental Rotation test (1978). The experimenter completed the experiment by debriefing the participant and answered all questions thoroughly. The entire session (including all paperwork) took approximately 120 min.

2.4 Experimental Design and Performance Measures

The experiment was a between subject design with the following as dependent variables: percentage correct on each SA probe, operator trust according to the respective scales, and perceived workload. Level of transparency information displayed was the independent variable. OSPAN, attentional control, video game efficacy, and mental rotation scores were used as covariates. For more information about the performance measures see Appendix I.

3. Results

The experiment contained several measures across the dimensions of SA and trust. Table 1 provides the means and standard deviations to examine results across experimental conditions.

Table 1 Summary of means and standard deviations according to transparency level

Measure	Group 1	Group 2	Group 3	Total
SA 1 – Which resources are currently green?	0.924 (0.106)	0.859 (0.115)	0.867 (0.106)	0.883 (0.11)
SA 2 – Which resources were last reduced?	0.793 (0.189)	0.869 (0.062)	0.763 (0.114)	0.809 (0.137)
SA 3 – Which resource does the ASM need to be least concerned about?	0.817 (0.226)	0.783 (0.269)	0.768 (0.152)	0.79 (0.217)
SA 4 – How many times have you stopped to answer questions?	0.633 (0.293)	0.728 (0.140)	0.682 (0.150)	0.681 (0.205)
SA 5 – When was the last time your current status changed?	0.595 (0.168)	0.627 (0.182)	0.643 (0.181)	0.622 (0.174)
SA 6 – How many hazard zones are currently visible?	0.608 (0.405)	0.884 (0.155)	0.891 (0.079)	0.794 (0.282)
SA 7 – What type of hazard did the ASM last go through?	0.371 (0.317)	0.790 (0.107)	0.806 (0.088)	0.656 (0.282)
SA 8 – Why were the resources reduced?	0.308 (0.317)	0.855 (0.129)	0.792 (0.165)	0.652 (0.327)
Trust – Modified Jian 1 pre	54.6 (11.12)	57.2 (9.58)	61.47 (9.21)	57.76 (10.18)
Trust – Modified Jian 1 post	52.67 (9.83)	60.67 (8.96)	60.07 (9.758)	57.8 (10.0)
Trust – Schaefer pre	65.64 (18.15)	75.26 (11.46)	73.79 (11.08)	71.56 (14.28)
Trust – Schaefer post	65.31 (20.56)	79.02 (13.3)	73.02 (12.55)	72.45 (16.52)
Trust – Modified Jian 2 gathering and filtering information	10.07 (8.51)	18.8 (5.0)	12.93 (8.64)	13.93 (8.26)
Trust – Modified Jian 2 integrating and displaying analyzed information	5.07 (11.07)	15.87 (7.02)	12.67 (7.39)	11.2 (9.65)
Trust – Modified Jian 2 suggesting or making decisions	–2.8 (3.0)	1.8 (2.7)	–1.47 (3.09)	–822 (3.47)
Trust – Modified Jian 2 executing actions	14.6 (4.53)	20.87 (4.45)	20.6 (5.5)	18.69 (5.57)
Workload – overall	38.08 (18.49)	37.16 (17.12)	41.38 (15.91)	38.87 (16.91)

Each SA probe was checked for violations of the assumptions. Exploration of the data indicated large violations of normality for the SA probes. These violations were confirmed via 3 different methods: graphing of a histogram with a normal curve, getting standardized skewness and kurtosis measures, and via the Shapiro-Wilk test. Transformations were attempted but were unsuccessful at correcting for normality.

The SA probes were then analyzed for correlations between each of the probes. Results showed moderate correlations between the probes (Tables 2 and 3). There is evidence in the literature for the analysis of variance (ANOVA) to be robust to normality violations (Norman 2010). While there is danger of Type II error or false negative (Fayers 2011), the Box's M test for homogeneity of covariance and Levene's test for equality of variances can be used to support the performance of a multivariate analysis of variance (MANOVA). We used these measures as a validation check, combined with reporting of effect sizes, to facilitate the selection of a MANOVA analysis of groups of SA questions. There were 2 outliers, which are scores that have Z-scores in excess of 3.29 according to Tabachnick and Fidell (2012), and the value was adjusted one unit away from the next extreme outlier.

Table 2 Correlations among situation awareness probes that can be determined by all groups

SA Probe	1	2	3	4	5
1 – Which resources are currently green?	…	…	…	…	…
2 – Which resources were last reduced?	0.591[a]	…	…	…	…
3 – Which resource should the ASM be least concerned about?	0.412[a]	0.531[a]	…	…	…
4 – How many times have you stopped during the route to answer questions?	0.260	0.496[a]	0.293	…	…
5 – When was the last time your current status icon changed?	–0.083	0.445[a]	0.520[a]	0.336[b]	…

[a]Correlation is significant at the 0.01 level (2-tailed).
[b]Correlation is significant at the 0.05 level (2-tailed).

Table 3 Correlations between situation awareness probes involving transparency conditions

SA Probe	1	2	3	4	5	6	7	8
1 – Which resources are currently green?	…	…	…	…	…	…	…	…
2 – Which resources were last reduced?	0.488[a]	…	…	…	…	…	…	…
3 – Which resource should the ASM be least concerned about?	0.361	0.399[a]	…	…	…	…	…	…
4 – How many times have you stopped during the route to answer questions?	–0.017	0.004	–0.113	…	…	…	…	…
5 – When was the last time your current status icon changed?	0.122	0.220	0.433[a]	–0.114	…	…	…	…
6 – How many hazard zones are currently visible?	0.322	0.247	0.017	0.250	0.125	…	…	…
7 – What hazards did the ASM last go through?	0.617[b]	0.285	0.435[a]	–0.019	0.201	0.411[a]	…	…
8 – Why was the resource reduced?	0.633[b]	0.613[b]	0.362[a]	0.111	0.330	0.427[a]	0.817[b]	…

[a]Correlation is significant at the 0.01 level (2-tailed).
[b]Correlation is significant at the 0.05 level (2-tailed).

Each probe was presented 3 times per scenario, over a total of 6 scenarios, totaling 18 instances of answering the question. The scoring for responses is a ratio scale of a number of correct choices selected/total number of correct choices. All questions had either 1, 2, or 3 correct answers. Participants received no credit for a wrong answer. The lowest score possible was 0%, and the highest was 100%. The average of 18 responses was used as the question score for the analysis.

Repeated measures ANOVAs were used to evaluate the effect of agent transparency information on trust according to a pre-post design of 2 different trust scales, $\alpha = 0.05$. A second modified trust questionnaire was administered postexperiment. This questionnaire was designed according to Parasuraman et al.'s (2000) levels of interacting with automated systems, $\alpha = 0.05$. Aggregate scores were created to allow comparisons between levels. Perceived workload, according to the instantaneous self-assessment (ISA), was measured using between subjects ANOVAs. For the NASA-TLX, a repeated measures ANOVA was used to evaluate the effects of agent transparency information on the perceived workload, $\alpha = 0.05$.

3.1 Situation Awareness

There are 2 sets of data analysis for SA: One set of statistical analysis compared SA probes across all conditions, and the second set compared the SA probes exclusive to the latter 2 transparency conditions. Analysis was performed on 2 sets because a few of the SA probes asked for information that was not displayed in group 1. Although the correct answer for information not displayed would have been "I don't know", after initial analysis the authors felt that this was unfair to the participants in group 1. Therefore, those questions were removed from the all group analysis.

3.1.1 Analysis Including All Groups

For this analysis, the SA probes significantly correlated with each other (Table 2). Additional correlation tables according to group can be found in Appendix J. There was not a clear trend of increasing correlations between groups. These significant correlations, coupled with examination of Box's M test, Levene's test, and effect sizes, fulfill some underlying prerequisites for MANOVA, which suggest that the results would accurately reflect the world.

Examination of the multivariate assumptions with all 5 questions included indicated violations of both Box's M test, $p < 0.001$, and one probe, "How many times have you stopped during the route to answer questions", indicated a violation of Levene's test, $p = 0.006$. Therefore, this question was removed from the MANOVA. The remaining 4 questions, complied with assumptions of normality tested by Box's M test, $p = 0.008$ and Levene's test, all p's > 0.05.

The combined dependent variables (DVs) were significantly affected by experimental condition, Wilks' Lambda = 0.623, $F(8,78) = 2.603$, $p = 0.014$. The results reflected a modest association between experimental conditions (group 1, $M = 0.782$; group 2, $M = 0.784$; group 3, $M = 0.760$) $\eta^2 = 0.37$. Since it was not possible to measure some SA probes for group 1, they were excluded. If these probes were included, the differences would have been even larger. To investigate the impact of experimental condition on the individual DVs, post hoc comparisons were conducted using the Bonferroni correction, but all results were nonsignificant. To investigate the effect of individual differences on the SA probes, attentional control, video game experience, both OSPAN scores, and mental rotation were analyzed separately as covariates. When incorporating the OSPAN math score, the model improved in significance, Wilks' Lambda = 0.607, $F(8,76) = 2.697$, $p = 0.011$, $\eta^2 = 0.39$. This is interesting especially when considering the size of the sample.

3.1.2 Analysis Including Only Group 2 and Group 3

As with the previous analysis, the SA probes indicated significant correlations between each other (Table 3). The significant correlations, coupled with examination of Box's M test, Levene's test, and effect sizes, led to the use of MANOVA as the analysis technique.

Examination of the multivariate assumptions with all 8 questions included indicated multicollinearity between probes 7 and 8, p = 0.817, therefore question 8 was excluded from the analysis. Although the assumption of Box's M test was met, $p >$ 0.001, one probe, "Which resources were last reduced", indicated a violation of Levene's test, p = 0.027. Therefore, this question was removed from the MANOVA. Once the question was removed, the remaining 6 questions met the assumptions of Box's M test, p = 0.008 and Levene's test, all p's > 0.05.

With the use of Wilks' criterion, the combined DVs were not significantly affected by experimental condition, Wilks' Lambda = 0.954, $F(6,23) = 0.185$, $p = 0.978$. Overall, the analyses indicate partial support for H1, as operator SA did increase according to level of transparency information, but differential effects occurred depending on the question. The first analysis using the questions applicable to all levels produced significant results. However, the additional questions, when examining only group 2 and group 3 did not.

3.2 Trust

Three different measures were taken to assess operator trust, each is described in the following sections.

3.2.1 Modified Trust in Automated Systems Scale 1

No significant outliers were present as measured by Z-Scores. Sphericity, according to Mauchly's test, was violated, $\varepsilon > 0.75$, so the Huynh-Feldt correction was used. A significant interaction between change in trust and experimental condition was found, Wilks' Lambda = 0.863, $F(2, 42) = 3.344$, $p = 0.045$, $\eta^2 = 0.137$ (Fig. 6). Pairwise comparisons, using the Bonferroni correction ($\alpha = 0.017$), did not indicate any significant differences between individual levels.

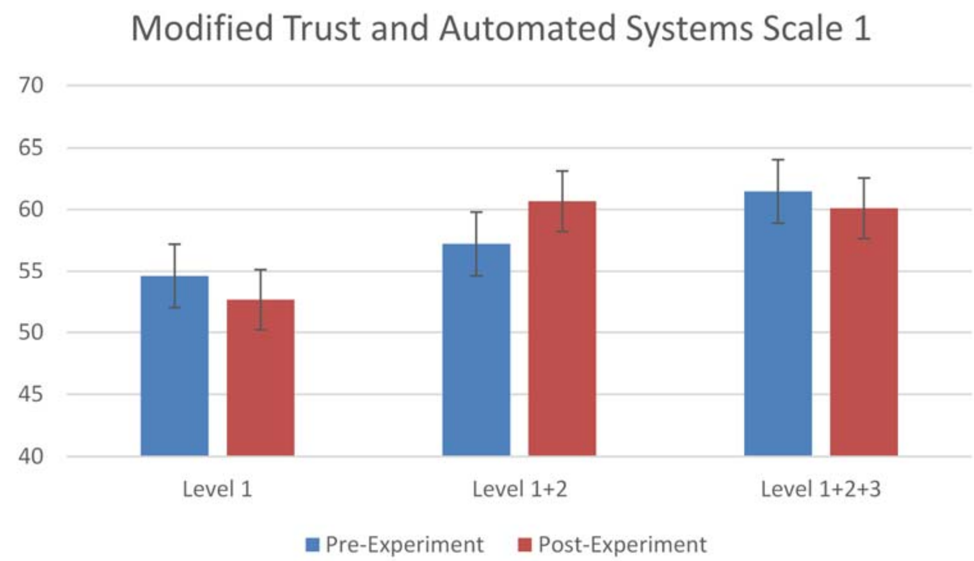


Fig. 6 Pre- and postresults of trust of automated systems scale 1 (error bars indicate standard error of the mean)

To attempt to reduce the error of the trust scores, the following individual differences were tested as covariates: attentional control, gaming experience, both OSPAN tests, and mental rotation. Three of the individual differences improved the interaction of scale score with experimental condition. Several covariates were tested for their effect on trust:

- Mental rotation: Wilks' Lambda = 0.845, $F(2, 41) = 3.747$, $p = 0.032$
- Gaming experience: Wilks' Lambda = 0.821, $F(2, 41) = 4.456$, $p = 0.018$
- Attentional control: Wilks' Lambda = 0.850, $F(2, 41) = 3.618$, $p = 0.036$

However, while the individual covariates helped explain the change in operator trust over time (the repeated measures variables), their addition did not make the differences in trust scores between transparency levels significant. Therefore, H4 was not supported.

3.2.2 Schaefer Human Robot Trust Scale

There was not a significant interaction between when the scale was administered and experimental condition when administered pre-post, Wilks' Lambda = 0.982, $F(2, 42) = 0.394$, $p = 0.677$, $\eta_p^2 = 0.018$.

3.2.3 Modified Trust in Automated Systems Scale 2

In examining the results of the scale, violations of normality were identified. These violations were confirmed via 3 different methods: graphing of a histogram with a normal curve, getting standardized skewness and kurtosis measures, and via the Shapiro-Wilk test. To allow for comparison of experimental conditions within the stages, aggregate variables of the scores were created. Questions where higher scores indicated higher trust were given positive values, while questions where higher scores indicated lower trust were given negative values. These values were combined and compared across experimental conditions (Fig. 7).

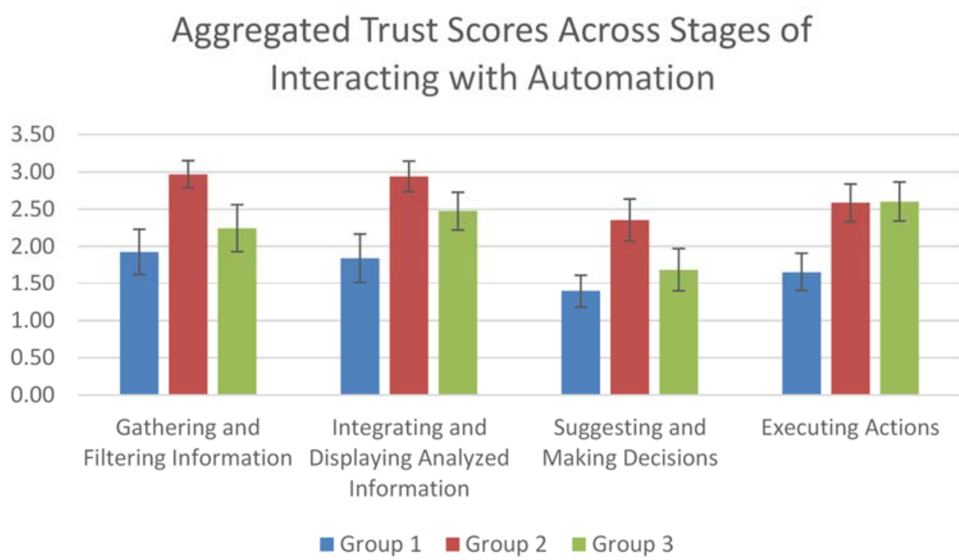


Fig. 7 Aggregate scores across stages of interacting with automation (error bars indicate standard error of the mean)

3.2.3.1 Gathering and Filtering Information

Participants in group 2 ($M = 2.97$, $SD = 0.71$) had the highest aggregate scores for this stage followed by participants in group 3 ($M = 2.25$, $SD = 1.23$), and participants in group 1 had the lowest ($M = 1.93$, $SD = 1.18$).

3.2.3.2 Integrating and Displaying Analyzed Information

Participants in group 2 ($M = 2.94$, $SD = 0.80$) had the highest scores for this stage followed by participants in group 3 ($M = 2.48$, $SD = 0.97$), and participants in group 1 had the lowest ($M = 1.84$, $SD = 1.27$).

3.2.3.3 Suggesting or Making Decisions

Participants in group 2 (*M* = 2.35, *SD* = 1.09) had the highest scores for this stage followed by participants in group 3 (*M* = 1.68, *SD* = 1.12), and participants in group 1 had the lowest (*M* = 1.40, *SD* = 0.84).

3.2.3.4 Executing Actions

Participants in group 3 (*M* = 2.60, *SD* = 1.02) had the highest scores for this stage followed by participants in group 2 (*M* = 2.59, *SD* = 0.98), and participants in group 1 had the lowest (*M* = 1.65, *SD* = 0.98).

Based on the results of the analysis, H2 was partially supported. Across the stages of automation, the transparency groups (2 and 3) consistently outperformed the baseline group. However, the performance between the 2 transparency groups is not different from each other.

3.3 Subjective Workload

In the assessment of subjective workload, 2 different measures were used: the NASA-TLX (Appendix K) and the ISA. The purpose of using 2 different measures was to investigate whether subjective workload would differ when workload was taken during the simulation (ISA) as opposed to after the simulation (NASA-TLX). There were 6 simulated scenarios, each with one ISA administration and one NASA-TLX administration. The ISA and the NASA-TLX were tested using reliability analyses to determine reliability of scores across scenarios. Both scales had extremely high reliability according to Cronbach's alpha (NASA-TLX = 0.98; ISA = 0.97); therefore, repeated measures ANOVA could be used for the NASA-TLX, but not for the ISA, as the data are noncontinuous. Analysis of the ISA data indicated large violations of normality via the Shapiro-Wilk test. Therefore, the Kruskal-Wallis test was used to analyze the results.

3.3.1 Instantaneous Workload Assessment

There was not a significant difference in operator workload across the 3 agent transparency conditions for any of the 6 scenarios. This indicates that the scenario did not interact with agent transparency as measured by the ISA score.

3.3.2. NASA Task Load Index

A 6 (subscale) × 3 (experimental condition) repeated measures ANOVA was used to evaluate differences. Sphericity, according to Mauchly's test, was violated, $\varepsilon < 0.75$, so the Greenhouse-Geisser correction was used. There was a nonsignificant

interaction between subscale and experimental condition, $F(6.204, 130.289) = 0.579$, $p = 0.752$, partial $\eta^2 = 0.051$.

3.4 Individual Difference Factors

Individual difference factors were examined using Kruskal-Wallis tests for differences between experimental conditions (Table 4). The reason for this examination is that if the experimental conditions were not significantly different, it suggests the groups themselves were similar for the individual difference categories. The fact that the results came out nonsignificant is viewed as a positive outcome.

Table 4 Results of the Kruskal-Wallis H test for individual difference factors

Individual Difference Factor	Chi-Square	Degrees of Freedom	Asymptotic Significance
Mental rotation	2.391	2	0.303
OSPAN math	0.345	2	0.842
OSPAN letter	1.459	2	0.482
Gaming experience	3.304	2	0.192
Attentional control	1.138	2	0.566

3.5 System Usability

A between subjects ANOVA was used to evaluate the effect of agent transparency information on system usability, $\alpha = 0.05$. Examining system usability according to experimental condition, group 2 ($M = 78.40$, $SD = 11.60$) had the highest usability score, followed by group 3 ($M = 70.40$, $SD = 18.05$) and group 1 ($M = 66.47$, $SD = 14.55$). There was not a significant difference between transparency information conditions on the system usability scale $F(2, 42) = 2.477$, $p = 0.096$, $\eta^2 = 0.10$. Based on the results of the analysis, H5 was not supported as system usability did not increase with additional agent transparency information.

As a follow-up to the analysis on usability, a qualitative analysis based on the following postexperiment question was examined: "Which object in the interface did you use/find useful?" The participants could answer more than one object. The individuals in groups 1 and 2 performed as expected, group 1 predominantly used the current status icon (14) and group 2 most predominantly used zone overlays (11). Group 3 predominantly used the current status icon (Table 5). Further analysis of any additional comments by group 3 indicated no comments related to either predicted status icons or current status icons.

Table 5 Participant responses by group: which interface object did you use/find useful?

Condition	ASM Indicator	Route Markers	Zone Overlays	Uncertainty Zones	Current Status Icon	Predicted Status Icon
Group 1	3	2	0	0	14	0
Group 2	1	0	11	0	5	0
Group 3	1	0	3	0	12	0

4. Discussion

In the current study, we investigated whether increasing the level of transparency information improved operator's comprehension, trust, and usability of an intelligent agent while assessing workload and accounting for individual differences. Transparency information did contribute to differences between conditions on SA probes; however, follow-up analysis, once accounting for homogeneity of variance, showed no significant differences between individual groups. Workload did not increase with the addition of transparency information nor was system usability affected according to condition. The lack of differences demonstrated that information can be added related to the reasoning of an intelligent agent without affecting understanding of the situation. However, the subjective questions yielded an unexpected result to be discussed later in this section.

Looking at trust according to the stages of interacting with automation further explained this relationship. Across all 4 stages, a similar pattern emerged. Participants in group 2 demonstrated the most trust of the interface, followed by participants in group 3. The differences between these 2 conditions were much smaller than either condition's differences with group 1. It is possible that these 2 conditions were viewed as very similar and therefore, had similar trust levels.

The analysis of subjective workload using the NASA-TLX showed differences between the effects of subscale according to scenario. A possible explanation could be that since the maps were always presented in the same order, the users felt that their level of mental workload and effort decreased as they gained more experience with the interface. This research design includes a primarily passive task, thus without ways of interacting with the interface it becomes challenging to establish individual differences. In discussing the relationship between spatial ability and passive UGV performance, Ophir-Arbelle et al. (2013) found that an operator's spatial ability was not a significant predictor of performance. Similarly in the current study, spatial ability was not a significant predictor of SA. In another study, Oron-Gilad et al. (2011) found no significant correlations between performance

and gaming experience for a passive task by dismounted Soldiers while they did find correlations for an active task. The findings of this study are consistent with these results.

It is worth re-examining the results of the subjective question, “Which interface object did you use/find useful?” Although groups 1 and 2 performed as expected, group 3 reverted back to baseline, relying on the current status display. This research is consistent with other work in preparation in our lab, which found that during a route-planning task, individuals with high amounts of information reverted back to the baseline. Future research could potentially provide more scaffolding and a different way of providing prediction information to make them less similar. The discrepancy then for trust results of groups 1 and 3 could be that participants equated more information with being more trustworthy than minimal information but less trustworthy than the display with only the information they felt they needed. More research into investigating mapping out of domain and information requirements for this type of experiment and adjusting the interface accordingly would be beneficial and potentially change the results.

The largest limitation for this study was the lack of an adequate sample size. With only 15 participants per group, the results of this study are better served as a pilot study for future work. Also choosing a within-subjects design rather than a between-subjects design could have potentially led to identifying more significant differences between groups due to an increase in power. However, even with the small sample, modest effect sizes were found, indicating the potential for significance with a larger sample.

5. Conclusion

Previous research examined interface design for unmanned aerial vehicles, supervising multiple agents, and ecological interface design for command and control (Bennett and Flach 2011; Chen and Barnes 2014; Kilgore and Voshell 2014). This study focused on bridging the gap of conveying understanding with intelligent ground teammates.

This research found that through using straightforward, easy-to-understand displays operator trust of an intelligent agent increased. This supported past research efforts, which demonstrated that explanations of an agent’s reasoning can improve understanding and provide appropriate expectations to a human teammate (Lee and See 2004; Beck et al. 2007; Chen et al. 2011). The unique contribution of this research effort was examining higher-level understanding of displays related to UGVs and trust.

The results also emphasized how proper use of display elements can increase understanding without decreasing performance. The significance of these results demonstrates the effectiveness of agent transparency even on passive interfaces. Future research could investigate the possibility of using a more diverse group of interface design techniques to further describe the relationship between operator trust and agent transparency.

6. References

Adams E. Fundamentals of game design. 3rd ed. San Francisco (CA): New Riders; 2013.

Aretz AJ, Wickens CD. The mental rotation of map displays. Human Performance. 1992;5(4):303–328.

Barnes MJ, Evans AW. Soldier-robot teams in future battlefields: an overview. In: Barnes MJ, Jentsch F, editors. Human-Robot Interactions in Future Military Operation. Farnham, Surrey (UK): Ashgate; 2010. p. 9–29.

Beck HP, Dzindolet MT, Pierce LG. Automation usage decisions: controlling intent and appraisal errors in a target detection task. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2007;49(3):429–437.

Bennett KB, Flach JM. Display and interface design: subtle science, exact art. 1st ed. Boca Raton (FL): CRC Press; 2011.

Brooke J. SUS-A quick and dirty usability scale. Usability Evaluation in Industry. 1996;189–194.

Chen JY, Barnes MJ. Human–agent teaming for multirobot control: a review of human factors issues. IEEE Transactions on Human-Machine Systems. 2014;44(1):13–29.

Chen JY, Barnes MJ, Harper-Sciarini M. Supervisory control of multiple robots: human-performance issues and user-interface design. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews. 2011;41(4):435–454.

Chen JY, Procci K, Boyce M, Wright J, Garcia A, Barnes M. Situation awareness-based agent transparency. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2014. Report No.: ARL-TR-6905. Also available at http://www.arl.army.mil/www/default.cfm?technical_report=7066.

Derryberry D, Reed MA. Anxiety-related attentional biases and their regulation by attentional control. Journal of Abnormal Psychology. 2002;111(2):225–236.

Endsley MR. Design and evaluation for situation awareness enhancement. In: Proceedings of the 32nd Human Factors and Ergonomics Society Annual Meeting; 1988 Oct 24–28; Anaheim, CA. Santa Monica (CA): Human Factors and Ergonomics Society; 1988. p. 97–101.

Endsley MR. Situation awareness. In: Salvendy G, editor. Handbook of human factors and ergonomics. 4th ed. Hoboken (NJ): Wiley; 2012. p. 553–568.

Endsley M, Jones D. Designing for situation awareness an approach to user-centered design. 2nd ed. Boca Raton (FL): CRC Press; 2012.

Engle RW. Working memory capacity as executive attention. Current Directions in Psychological Science. 2002;11(1):19–23.

Fayers P. Alphas, betas and skewy distributions: two ways of getting the wrong answer. Advances in Health Sciences Education. 2011;16(3):291–296.

Hancock PA, Billings DR, Schaefer KE, Chen JY, De Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2011;53(5):517–527.

Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. Advances in Psychology. 1988;52:139–183.

Hoffman RR, Johnson M, Bradshaw JM, Underbrink A. Trust in automation. Intelligent Systems, IEEE. 2013;28(1):84–88.

Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics. 2000;4(1):53–71.

Jordan CS, Brennen SD. Instantaneous self-assessment of workload technique (ISA). [place unknown]: Eurocontrol SKYbrary Bookshelf; 1992 [accessed 2015 Apr 20]. http://www.skybrary.aero/bookshelf/books/1963.pdf.

Kilgore R, Voshell M. Increasing the transparency of unmanned systems: applications of ecological interface design. In: Shumaker R, Lackey S, editors. Virtual, augmented and mixed reality: applications of virtual and augmented reality. VAMR 2014. Proceedings of the 6th International Conference, Part II; 2014 Jun 22–27; Heraklion, Crete, Greece. New York (NY): Springer; 2014. p. 378–389.

Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. Ergonomics. 1992;35(10):1243–1270.

Lee JD, See KA. Trust in automation: designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2004;46(1):50–80.

Lyons JB. Being transparent about transparency: a model for human-robot interaction. Papers from the 2013 AAAI Spring Symposium Series. Palo Alto (CA): The AAAI Press; 2013. Technical Report SS-13-07.

Lyons JB, Havig PR. Transparency in a human-machine context: approaches for fostering shared awareness/intent. In: Shumaker R, Lackey S, editors. Virtual, augmented and mixed reality: designing and developing virtual and augmented environments. VAMR 2014. Proceedings of the 6th International Conference, Part I; 2014 Jun 22–27; Heraklion, Crete, Greece. New York (NY): Springer; 2014. p. 181–190.

Miller, CA. Delegation and transparency: coordinating interactions so information exchange is no surprise. In: Shumaker R, Lackey S, editors. Virtual, augmented and mixed reality: designing and developing virtual and augmented environments VAMR 2014. Proceedings of the 6th International Conference, Part I; 2014 Jun 22–27; Heraklion, Crete, Greece. New York (NY): Springer; 2014. p. 191–202.

Norman G. Likert scales, levels of measurement and the laws of statistics. Advances in Health Sciences Education. 2010;15(5):625–632.

Ophir-Arbelle R, Oron-Gilad T, Borowsky A, Parmet Y. Is more information better? How dismounted Soldiers utilize video feed from unmanned vehicles: attention allocation and information extraction considerations. Journal of Cognitive Ergonomics and Decision Making. 2013;7(1):26–48.

Oron-Gilad T, Redden ES, Minkov Y. Robotic displays for dismounted warfighters a field study. Journal of Cognitive Engineering and Decision Making. 2011;5(1):29–54.

Ososky S, Sanders T, Jentsch F, Hancock P, Chen JYC. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. Proc. SPIE 9084, Unmanned Systems Technology XVI, 2014 Jun 3. doi: 10.1117/12.2050622.

Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. 2000;30(3):286–297.

Rao AS, Georgeff MP. BDI agents: from theory to practice. ICMAS. 1995;95:312–319.

Rasmussen J, Vicente KJ. Coping with human errors through system design: implications for ecological interface design. International Journal of Man-Machine Studies. 1989;31(5):517–534.

Rasmussen J, Vicente, KJ. Ecological interfaces: A technological imperative in high-tech systems? International Journal of Human-Computer Interaction. 1990:2(2):93110. http://dx.doi.org/10.1080/10447319009525974.

Rehfeld, S. The impact of mental transformation training across levels of automation on spatial awareness in human-robot interaction [dissertation]. [Orlando (FL)]: University of Central Florida; 2006.

Rouse WB, Morris NM. On looking into the black box: prospects and limits in the search for mental models. Atlanta (GA): Georgia Institute of Technology Atlanta Center for Human-Machine Systems Research; 1985. Report No. (OTIC AD-A159080).

Russell S, Norvig P. Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River (NJ): Prentice Hall; 2009.

Schaefer KE. The perception and measurement of human-robot trust [dissertation]. [Orlando (FL)]: University of Central Florida; 2013.

Tabachnick BG, Fidell LS. Using multivariate statistics. 6th ed. Boston (MA): Pearson Education; 2012.

Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills. 1978;47(2):599–604.

Appendix A. Demographic Questionnaire

This appendix appears in its original form, without editorial change.

Participant # _______ Age ______ Major ______________Date __________ Gender ___

1. What is the <u>highest</u> level of education you have had?

   Less than 4 yrs of college ____ Completed 4 yrs of college ____ Other ____

2. When did you use computers in your education? (*<u>Circle all that apply</u>*)

   Grade School Jr. High High School Technical School College
   Did Not Use

3. Where do you currently use a computer? (*<u>Circle all that apply</u>*)

   Home Work Library Other________ Do Not Use

4. For each of the following questions, <u>circle</u> the response that best describes you.

   How often do you:

   Use a mouse? Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use a joystick? Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use a touch screen? Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use icon-based programs/software?
   Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use programs/software with pull-down menus?
   Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use graphics/drawing features in software packages?
   Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Use E-mail? Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Operate a radio controlled vehicle (car, boat, or plane)?
   Daily, Weekly, Monthly, Once every few months, Rarely, Never
   Play computer/video games?
   Daily, Weekly, Monthly, Once every few months, Rarely, Never

5. Which type(s) of computer/video games do you most often play if you play at least once every few months?

6. Which of the following best describes your expertise with computer? (check √ one)

   _____ Novice
   _____ Good with one type of software package (such as word processing or slides)
   _____ Good with several software packages
   _____ Can program in one language and use several software packages
   _____ Can program in several languages and use several software packages

7. Are you in your good/ comfortable state of health physically? YES NO

If NO, please briefly explain:

8. How many hours of sleep did you get last night? ______ hours

9. Do you have normal color vision? YES NO

10. Do you have military service? YES NO If Yes, how long __________

INTENTIONALLY LEFT BLANK.

Appendix B. Vandenberg and Kuse Mental Rotation Test

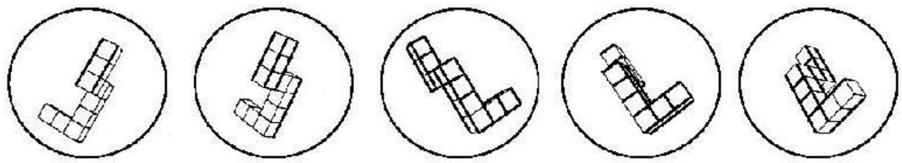
This appendix appears in its original form, without editorial change.

Name ____________________

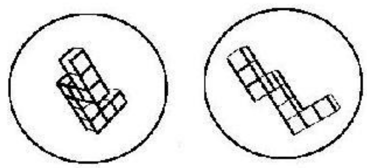
M.R.T. Test | Date ____________________

____

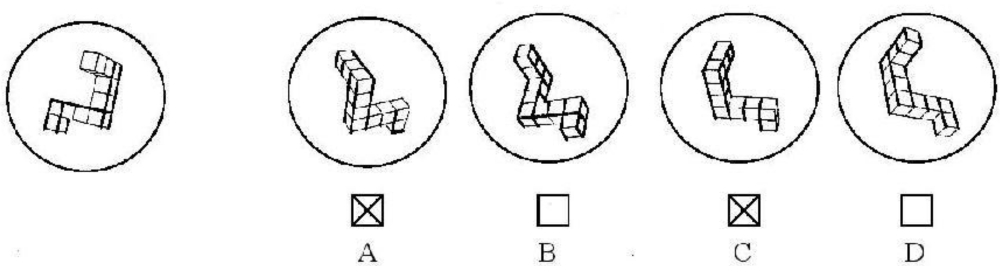
This is a test of your ability to look at a drawing of a given object and find the same object within a set of dissimilar objects. The only difference between the original objects and the chosen object will be that they are presented at different angles. An illustration of this principle is given below, where the same single object is given in five different positions. Look at each of them to satisfy yourself that they are only presented at different angles from one another.



Below are two drawings of new objects. They cannot be made to match the above five drawings. Please note that you may not turn over the objects. Satisfy yourself that they are different from the above.



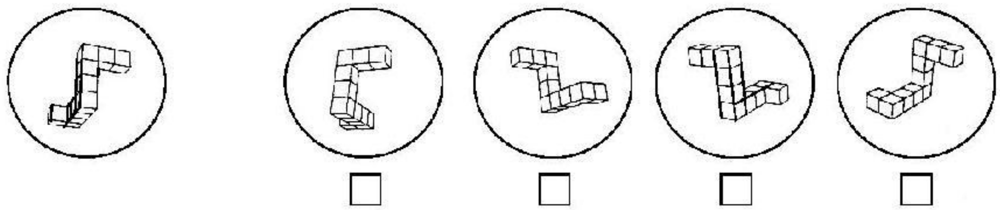
Now let's do some sample problems. For each problem there is a primary object on the far left. You are to determine which two of four objects to the right are the same object given on the far left. In each problem always <u>two</u> of the four drawings are the same object as the one on the left. You are to put Xs in the boxes below the correct ones, and leave the incorrect ones blank. The first sample problem is done for you.



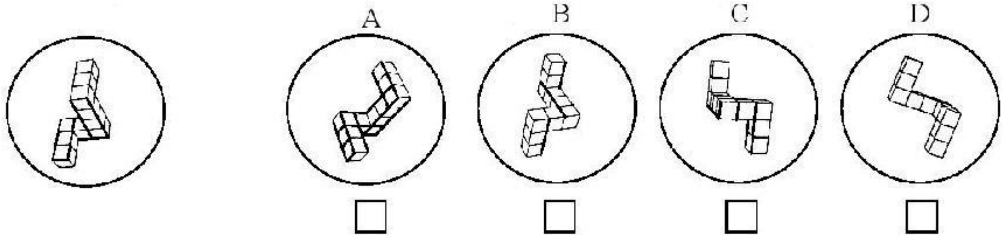
Go to the next page

Adapted by S.G. Vandenberg, University of Colorado, July 15, 1971
Revised instructions by H. Crawford, U. of Wyoming, September, 1979
Digitally remastered by S. Rehfeld and S. Scielzo, U. of Central Florida, July 2005

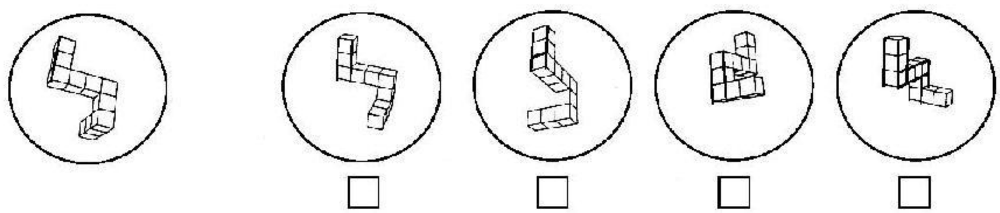
Do the rest of the sample problems yourself. Which two drawings of the four on the right show the same objects as the one on the left? There are always two and only two correct answers for each problem. Put an X under the two correct drawings.



A B C D



A B C D



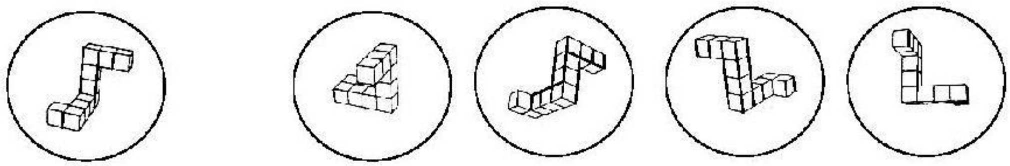
A B C D

Answer: (1) first and second drawing are correct
(2) first and third drawing are correct
(3) second and third drawing are correct

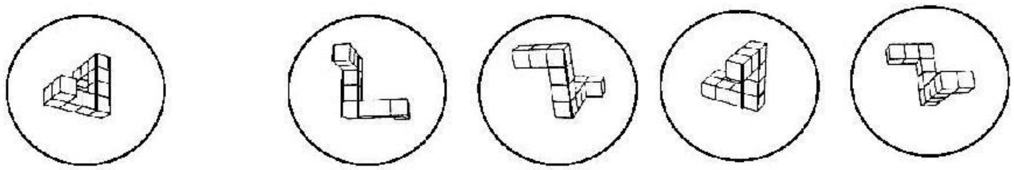
This test has two parts. You will have 3 minutes for each of the two parts. Each part has two pages. When you have finished Part I, STOP. Please do not go on to Part 2 until you are asked to do so. Remember: There are always two and only two correct answers for each item.

Work as quickly as you can without sacrificing accuracy. Your score on this test will reflect both the correct and incorrect responses. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct.

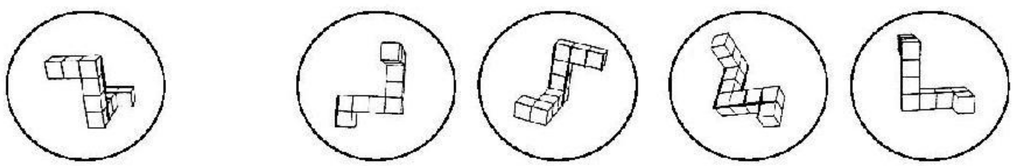
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

1. 

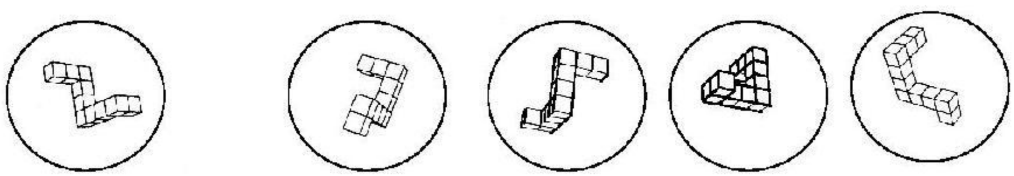
A B C D

2. 

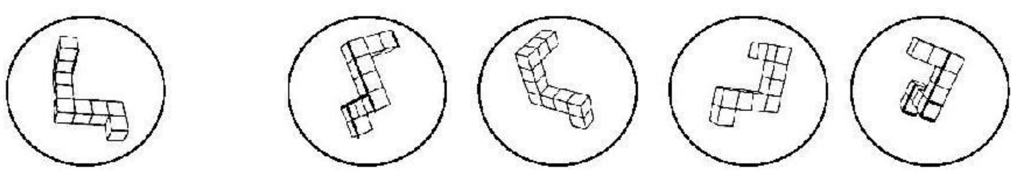
A B C D

3. 

A B C D

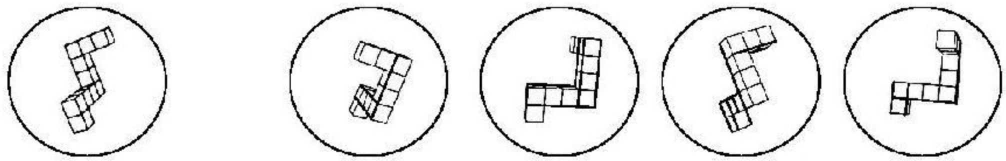
4. 

A B C D

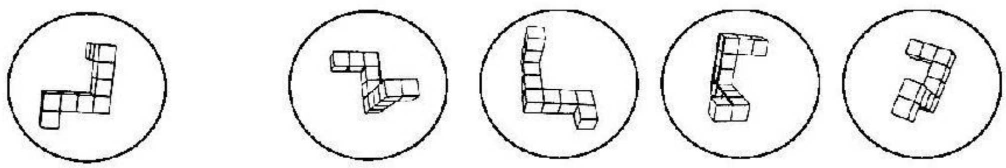
5. 

A B C D

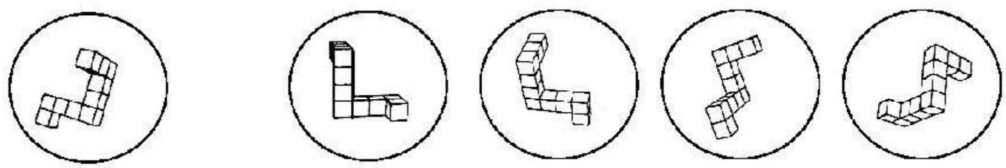
GO ON TO THE NEXT PAGE

6. 

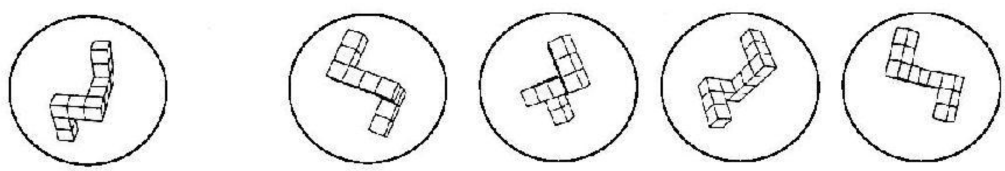
A B C D

7. 

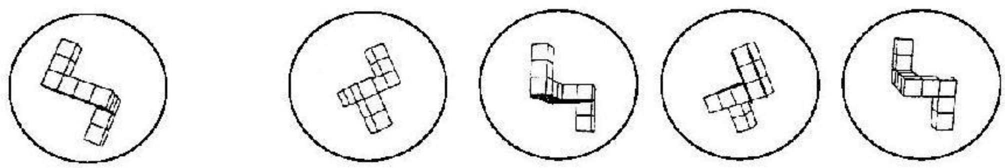
A B C D

8. 

A B C D

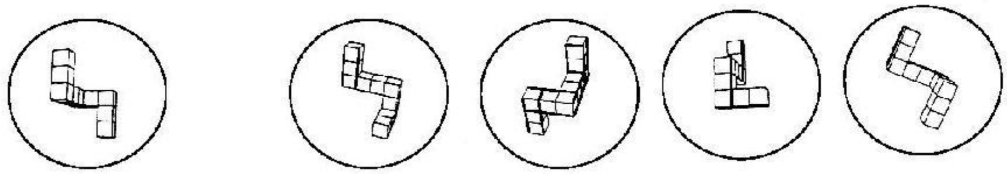
9. 

A B C D

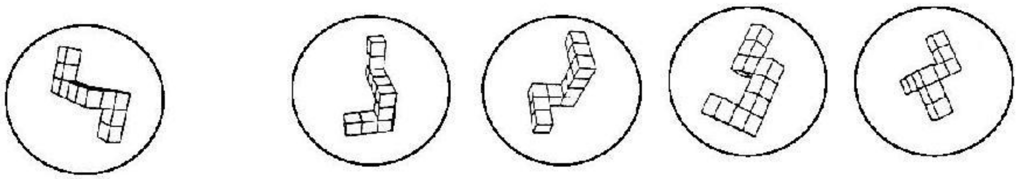
10. 

A B C D

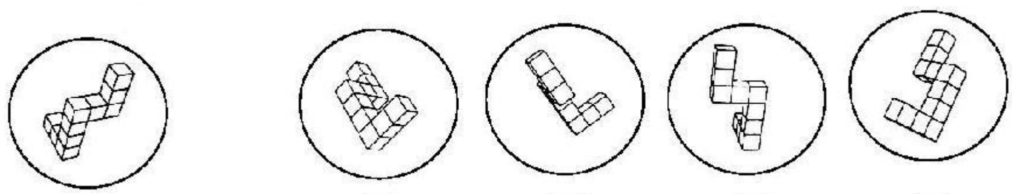
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO STOP

11. 

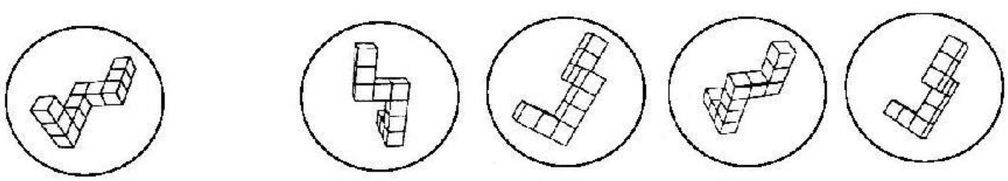
A B C D

12. 

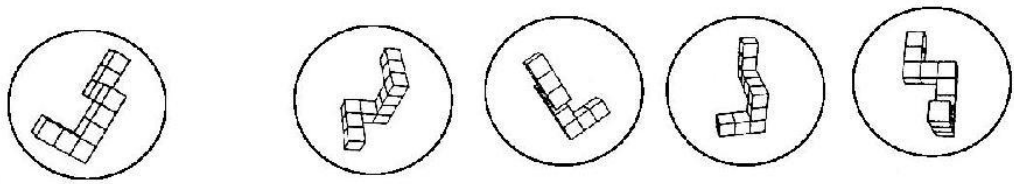
A B C D

13. 

A B C D

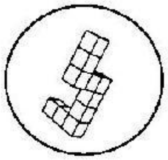
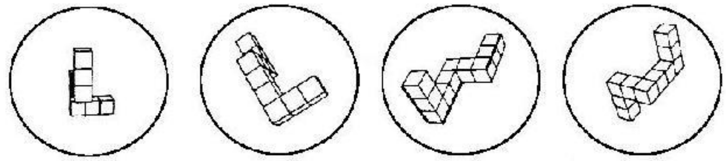
14. 

A B C D

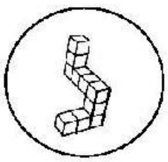
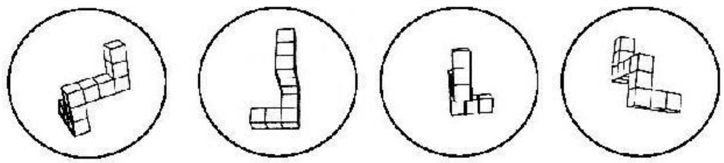
15. 

A B C D

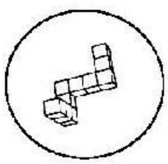
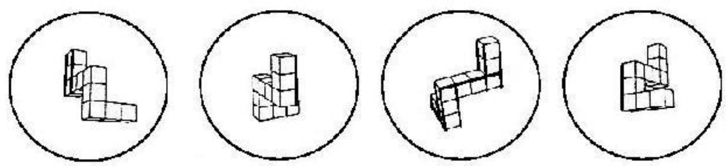
GO ON TO THE NEXT PAGE

16.  

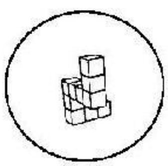
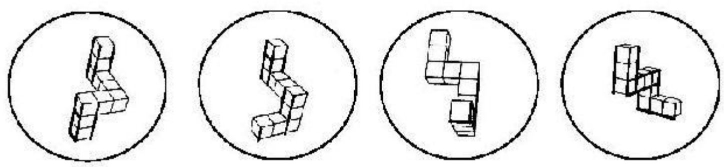
☐ A ☐ B ☐ C ☐ D

17.  

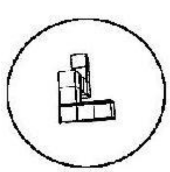
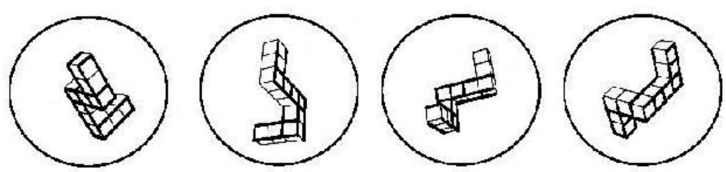
☐ A ☐ B ☐ C ☐ D

18.  

☐ A ☐ B ☐ C ☐ D

19.  

☐ A ☐ B ☐ C ☐ D

20.  

☐ A ☐ B ☐ C ☐ D

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO STOP

INTENTIONALLY LEFT BLANK.

Appendix C. Attentional Control Survey

This appendix appears in its original form, without editorial change.

Participant #_______ **Date _________**

For each of the following questions, circle the response that best describes you.

It is very hard for me to concentrate on a difficult task when there are noises around.

Almost never, Sometimes, Often, Always

When I need to concentrate and solve a problem, I have trouble focusing my attention.

Almost never, Sometimes, Often, Always

When I am working hard on something, I still get distracted by events around me.

Almost never, Sometimes, Often, Always

My concentration is good even if there is music in the room around me.

Almost never, Sometimes, Often, Always

When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me. Almost never, Sometimes, Often, Always

When I am reading or studying, I am easily distracted if there are people talking in the same room.

Almost never, Sometimes, Often, Always

When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.

Almost never, Sometimes, Often, Always

I have a hard time concentrating when I'm excited about something.

Almost never, Sometimes, Often, Always

When concentrating, I ignore feelings of hunger or thirst.

Almost never, Sometimes, Often, Always

I can quickly switch from one task to another.

Almost never, Sometimes, Often, Always

It takes me a while to get really involved in a new task.

Almost never, Sometimes, Often, Always

It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.

Almost never, Sometimes, Often, Always

I can become interested in a new topic very quickly when I need to.
Almost never, Sometimes, Often, Always

It is easy for me to read or write while I'm also talking on the phone.
Almost never, Sometimes, Often, Always

I have trouble carrying on two conversations at once.
Almost never, Sometimes, Often, Always

I have a hard time coming up with new ideas quickly.
Almost never, Sometimes, Often, Always

After being interrupted or distracted, I can easily shift my attention back to what I was doing before.
Almost never, Sometimes, Often, Always

When a distracting thought comes to mind, it is easy for me to shift my attention away from it.
Almost never, Sometimes, Often, Always

It is easy for me to alternate between two different tasks.
Almost never, Sometimes, Often, Always

It is hard for me to break from one way of thinking about something and look at it from another point of view.
Almost never, Sometimes, Often, Always

INTENTIONALLY LEFT BLANK.

Appendix D. System Usability Scale

This appendix appears in its original form, without editorial change.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system					

Appendix E. Modified Jian Pre-Post Trust Survey

This appendix appears in its original form, without editorial change.

Automation Survey

Automation refers to a system that reduces the need for human work. According to Lee and See (2004), “Automation is technology that actively selects data, transforms information, makes decisions, or controls processes.” Below is a statement evaluating your feelings about automation. Please circle the number that best describes your feeling or impression.

1 = not at all; 7 = extremely

1. Automation is deceptive.

1 2 3 4 5 6 7

2. Automation systems behave in an underhanded manner.

1 2 3 4 5 6 7

3. I am suspicious of the intent, action, or outputs of automation.

1 2 3 4 5 6 7

4. I am wary of automation.

1 2 3 4 5 6 7

5. The actions of automated systems will have harmful or injurious outcomes.

1 2 3 4 5 6 7

6. I am confident in automation.

1 2 3 4 5 6 7

7. Automated systems provide security.

1 2 3 4 5 6 7

8. Automated systems have integrity.

1 2 3 4 5 6 7

9. Automated systems are dependable.

1 2 3 4 5 6 7

10. Automated systems are reliable.

1 2 3 4 5 6 7

11. I can trust automated systems.

1 2 3 4 5 6 7

The Trust Survey is based on the questionnaire of Human-Computer Trust from Jian et al. (1998)

Appendix F. Posttest Modified Jian Trust Survey 2

This appendix appears in its original form, without editorial change.

For each of the following items and situations, circle the number which best describes your feeling or your impression based on the system you just used. For each item, consider the following situations:

- A: When the system is collecting and/or highlighting/filtering information.
- B: When the system is integrating information, generating predictive displays, and/or presenting its analysis.
- C: When the system is making decisions and/or selecting actions.
- D: When the system is executing actions.

1. The system is deceptive when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

2. The system behaves in an underhanded manner when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

3. I am suspicious of the system's intent, action, or outputs when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

4. I am wary of the system when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

5. The system's actions will have a harmful or injurious outcome when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

6. I am confident in the system when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

7. The system provides security when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

8. The system has integrity when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

9. The system is dependable when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

10. The system is reliable when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

11. I can trust the system when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

12. I am familiar with the system when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

13. The system is predictable when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

14. The system meets the needs of the mission when…

	not at all			*neutral*		*extremely*	
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

15. The system provides appropriate information when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

16. The system malfunctions when…

	not at all			*neutral*			*extremely*
A: Gathering or Filtering Information	1	2	3	4	5	6	7
B: Integrating and Displaying Analyzed Information	1	2	3	4	5	6	7
C: Suggesting or Making Decisions	1	2	3	4	5	6	7
D: Executing Actions	1	2	3	4	5	6	7

Now imagine that you are employed as an unmanned vehicle operator to complete missions. Reflecting on the experience with the system you just used, please rate the extent to which you agree with each of these items by circling a value from **1** (**strongly disagree**) to **7** (**strongly agree**), where **4** is **neutral**.

	Strongly Disagree			**Neutral**			**Strongly Agree**
17. Using the system would improve my job performance.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
18. Using the system would make it easier to do my job.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
19. I would find the system useful in my job.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
20. Learning to operate the system is easy for me.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
21. It is easy for me to become skillful at using the system.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
22. I find the system easy to use.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
23. I intend to use this system for my job.	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Appendix G. Schaefer Pre Trust Survey

This appendix appears in its original form, without editorial change.

PRE-SCALE

PARTCIPANT # ______________________________ (Page 1 of 3)

Now that you have seen a picture of the robot you will be working with, please rate the following items about this robot.

What % of the time will this robot be…	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1 Function successfully	O	O	O	O	O	O	O	O	O	O	O
2 Act consistently	O	O	O	O	O	O	O	O	O	O	O
3 Reliable	O	O	O	O	O	O	O	O	O	O	O
4 Predictable	O	O	O	O	O	O	O	O	O	O	O
5 Dependable	O	O	O	O	O	O	O	O	O	O	O
6 Follow directions	O	O	O	O	O	O	O	O	O	O	O
7 Meet the needs of the mission	O	O	O	O	O	O	O	O	O	O	O
8 Perform exactly as instructed	O	O	O	O	O	O	O	O	O	O	O
9 Have errors	O	O	O	O	O	O	O	O	O	O	O
10 Provide appropriate information	O	O	O	O	O	O	O	O	O	O	O
11 Unresponsive	O	O	O	O	O	O	O	O	O	O	O
12 Malfunction	O	O	O	O	O	O	O	O	O	O	O
13 Communicate with people	O	O	O	O	O	O	O	O	O	O	O
14 Provide feedback	O	O	O	O	O	O	O	O	O	O	O

Appendix H. Schaefer Post Trust Survey

This appendix appears in its original form, without editorial change.

POST-SCALE

PARTCIPANT # ________________________________ (Page 1 of 3)

Now that you have interacted with the robot, please rate the following items about this robot.

What % of the time will this robot be…	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1 Function successfully	O	O	O	O	O	O	O	O	O	O	O
2 Act consistently	O	O	O	O	O	O	O	O	O	O	O
3 Reliable	O	O	O	O	O	O	O	O	O	O	O
4 Predictable	O	O	O	O	O	O	O	O	O	O	O
5 Dependable	O	O	O	O	O	O	O	O	O	O	O
6 Follow directions	O	O	O	O	O	O	O	O	O	O	O
7 Meet the needs of the mission	O	O	O	O	O	O	O	O	O	O	O
8 Perform exactly as instructed	O	O	O	O	O	O	O	O	O	O	O
9 Have errors	O	O	O	O	O	O	O	O	O	O	O
10 Provide appropriate information	O	O	O	O	O	O	O	O	O	O	O
11 Unresponsive	O	O	O	O	O	O	O	O	O	O	O
12 Malfunction	O	O	O	O	O	O	O	O	O	O	O
13 Communicate with people	O	O	O	O	O	O	O	O	O	O	O
14 Provide feedback	O	O	O	O	O	O	O	O	O	O	O

Appendix I. More Information on Performance Measures

This appendix appears in its original form, without editorial change.

Situation Awareness

SA is the perception and comprehension of the current state, reasoning, and projection of elements in the environment (Endsley 1995). The SA probe questions are:

1. Which Resources are Currently Green?
2. Which Resources Were Last Reduced?
3. Which Resource Does the Autonomous Squad Member Need to be Least Concerned About?
4. How Many Times Have you Stopped During the Route to Answer Questions?
5. When was the Last Time Your Current Status Icon Changed?
6. How Many Hazard Zones are Currently Visible?
7. What Type of Hazard did the ASM Last go Through?
8. Why Were the Resources Reduced?

Trust

Participants were given the Trust in Automated Systems scale (Jian et al. 2000) prior to the observation of the autonomous agent to establish a baseline of their trust in automation. The Trust in Automated Systems scale is a series of Likert scale items, ranging from 1 - 7 (1 = Not at all, 7 = Extremely). The questions encompassing the scale are:

1. The system is deceptive
2. The system behaves in an underhanded manner
3. I am suspicious of the system's intent, action, or outputs

4. I am wary of the system
5. The system's actions will have a harmful or injurious outcome
6. I am confident in the system
7. The system provides security
8. The system has integrity
9. The system is dependable
10. The system is reliable
11. I can trust the system
12. I am familiar with the system

They were also given the Schaefer (2013) scale on human-robot trust. The scale consists of 14 questions, where participants are asked to rate the robot from 0-100, based on the percentage of time the robot will act in the specified manner. At the start of the experiment, the participant views a picture of the robot then takes the pre-scale. The experimenter re-administers the scale after the experiment to assess the change in robot trust due to experimental manipulation.

The 14 questions that encompasses the scale are:

1. Function successfully
2. Act consistently
3. Reliable
4. Predictable
5. Dependable
6. Follow Directions
7. Meet the needs of the mission

8. Perform exactly as instructed
9. Have errors
10. Provide appropriate information
11. Unresponsive
12. Malfunction
13. Communicate with people
14. Provide feedback

Participants also rate their trust in the agent on the modified trust in automation scale. The modified scale assesses trust of the system as it corresponds with the four stages of human information processing (Parasuraman et al. 2000). The four stages include information acquisition, information analysis, decision and action selection, and action implementation. These stages were conceptualized in the scale as gathering or filtering information, integrating and displaying analyzed information, suggesting or making decisions, and executing actions. The modified scale included 16 questions, each scored on a 1-7 Likert scale, each of which asked about the four information processing automations. The 16 questions were:

1. The system is deceptive when
2. The system behaves in an underhanded manner when
3. I am suspicious of the system's intent, action, or outputs when
4. I am wary of the system when
5. The system's actions will have a harmful or injurious outcome when
6. I am confident in the system when

7. The system provides security when
8. The system has integrity when
9. The system is dependable when
10. The system is reliable when
11. I can trust the system when
12. I am familiar with the system when
13. The system is predictable when
14. The system meets the needs of the mission when
15. The system provides appropriate information when
16. The system malfunctions when

All three of the scales were measured using the participant's average scores. For the trust in automated system scales and the human robot trust scale, questions were scored as a group because of the survey design. For the modified scale, question scoring occured at three different levels:

1. Overall scale score
2. Aggregate score by question
3. Individual scores for each of the four subscales for each question.

Workload

Workload was assessed using two different measures:

1. ISA (Jordan and Brennan 1992). The ISA provides a measure of workload as the participants are in the middle of the experiment. The assessment asks the participant to rate the level of current workload 1-5 (1 = not at all, 5 = extremely). The workload prompt appears once per scenario.

2. NASA-TLX (Hart and Staveland 1988). The NASA-TLX is a validated assessment workload assessment measure used specifically for human-machine interaction. The measure has a series of subscales and relationships between different domains to determine an overall score (0-100, weighted). The subscales rate six different workloads: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration.

Appendix J. Correlation Tables for All SA Probes by Level

This appendix appears in its original form, without editorial change.

Group 1 Only:

SA Probe	1	2	3	4	5	6	7	8
1 – Which resources are currently green?	—							
2 – Which resources were last reduced?	.878**	—						
3 – Which resource should the ASM be least concerned about?	.491	.740**	—					
4 – How many times have you stopped during the route to answer questions?	.717**	.769**	.761**	—				
5 – When was the last time your current status icon changed?	.579*	.763**	.754**	.853**	—			
6 – How many hazard zones are currently visible?	-.383	-.186	.049	-.129	.025	—		
7 – What hazards did the ASM last go through?	-.576*	-.393	-.220	-.362	-.214	.616*	—	
8 – Why was the resource reduced?	-.504	-.372	-.280	-.367	-.254	.557*	.941**	—

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Group 2 Only:

SA Probe	1	2	3	4	5	6	7	8
1 – Which resources are currently green?	—							
2 – Which resources were last reduced?	.573*	—						
3 – Which resource should the ASM be least concerned about?	.523*	.638*	—					
4 – How many times have you stopped during the route to answer questions?	-.045	-.379	-.485	—				
5 – When was the last time your current status icon changed?	.040	.355	.482	-.143	—			
6 – How many hazard zones are currently visible?	.190	.236	.082	.400	.230	—		
7 – What hazards did the ASM last go through?	.894**	.508	.634	-.099	.258	.339	—	
8 – Why was the resource reduced?	.899**	.476	.445	.065	.269	.409	.947**	—

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Group 3 Only:

SA Probe	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1 – Which resources are currently green?	—							
2 – Which resources were last reduced?	.652*	—						
3 – Which resource should the ASM be least concerned about?	.083	.412	—					
4 – How many times have you stopped during the route to answer questions?	.022	.046	.466	—				
5 – When was the last time your current status icon changed?	.210	.264	.398	-.077	—			
6 – How many hazard zones are currently visible?	.641*	.518*	.202	.047	.068	—		
7 – What hazards did the ASM last go through?	.251	.353	.040	.100	.125	.613*	—	
8 – Why was the resource reduced?	.467	.664**	.332	.089	.416	.614*	.815**	—

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

INTENTIONALLY LEFT BLANK.

Appendix K. NASA-TLX Questionnaire

This appendix appears in its original form, without editorial change.

Please rate your <u>overall</u> impression of demands imposed on you during the exercise.

1. Mental Demand: How much mental and perceptual activity was required (e.g., thinking, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

2. Physical Demand: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

3. Temporal Demand: How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

4. Level of Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

5. Level of Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

6. Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

LOW |---|---|---|---|---|---|---|---|---| HIGH
1 2 3 4 5 6 7 8 9 10

Pairwise Comparison of Factors

Select the member of each pair that provided the most significant source of workload variation in these tasks.

Physical Demand vs. Mental Demand

Temporal Demand vs. Mental Demand

Performance vs. Mental Demand

Frustration vs. Mental Demand

Effort vs. Mental Demand

Temporal Demand vs. Physical Demand

Performance vs. Physical Demand

Frustration vs. Physical Demand

Effort vs. Physical Demand

Temporal Demand vs. Performance

Temporal Demand vs. Frustration

Temporal Demand vs. Effort

Performance vs. Frustration

Performance vs. Effort

Effort vs. Frustration

INTENTIONALLY LEFT BLANK.

List of Symbols, Abbreviations, and Acronyms

ANOVA	analysis of variance
ASM	autonomous squad member
DV	dependent variable
EID	ecological interface design
H	hypothesis
ISA	instantaneous self-assessment
MANOVA	multivariate analysis of variance
NASA-TLX	National Aeronautics and Space Administration-task load index
OSPAN	Operational Span
PAC	Perceived Attentional Control
SA	situation awareness
SAT	SA-based agent transparency model
SRK	symbol, rule, and knowledge
UGV	unmanned ground vehicle

1 (PDF) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA

2 (PDF) DIRECTOR
US ARMY RESEARCH LAB
RDRL CIO LL
IMAL HRA MAIL & RECORDS MGMT

1 (PDF) GOVT PRINTG OFC
A MALHOTRA

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM D
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL 35898-7290

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRS EA DR V J RICE
BLDG 4011 RM 217
1750 GREELEY RD
FORT SAM HOUSTON TX 78234-5002

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM DG
J RUBINSTEIN
BLDG 333
PICATINNY ARSENAL NJ 07806-5000

1 (PDF) ARMY RSCH LABORATORY – HRED
ARMC FIELD ELEMENT
RDRL HRM CH C BURNS
THIRD AVE BLDG 1467B
RM 336
FORT KNOX KY 40121

1 (PDF) ARMY RSCH LABORATORY – HRED
AWC FIELD ELEMENT
RDRL HRM DJ D DURBIN
BLDG 4506 (DCD) RM 107
FORT RUCKER AL 36362-5000

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM CK J REINHART
10125 KINGMAN RD
BLDG 317
FORT BELVOIR VA 22060-5828

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM AY M BARNES
2520 HEALY AVE
STE 1172 BLDG 51005
FORT HUACHUCA AZ 85613-7069

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM AP
D UNGVARSKY
POPE HALL BLDG 470
BCBL 806 HARRISON DR
FORT LEAVENWORTH KS 66027-2302

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM AR J CHEN
12423 RESEARCH PKWY
ORLANDO FL 32826-3276

1 (PDF) ARMY RSCH LAB – HRED
HUMAN SYSTEMS
INTEGRATION ENGR
TACOM FIELD ELEMENT
RDRL HRM CU P MUNYA
6501 E 11 MILE RD
MS 284 BLDG 200A
WARREN MI 48397-5000

1 (PDF) ARMY RSCH LABORATORY – HRED
FIRES CTR OF EXCELLENCE
FIELD ELEMENT
RDRL HRM AF
C HERNANDEZ
3040 NW AUSTIN RD RM 221
FORT SILL OK 73503-9043

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM AV
W CULBERTSON
91012 STATION AVE
FORT HOOD TX 76544-5073

1 (PDF) ARMY RSCH LABORATORY – HRED
RDRL HRM DE A MARES
1733 PLEASONTON RD BOX 3
FORT BLISS TX 79916-6816

8 (PDF) ARMY RSCH LABORATORY – HRED
SIMULATION & TRAINING TECHNOLOGY CENTER
RDRL HRT COL G LAASE
RDRL HRT I MARTINEZ
RDRL HRT T R SOTTILARE
RDRL HRT B N FINKELSTEIN
RDRL HRT G A RODRIGUEZ
RDRL HRT I J HART
RDRL HRT M C METEVIER
RDRL HRT S B PETTIT
12423 RESEARCH PARKWAY
ORLANDO FL 32826

1 (PDF) ARMY RSCH LABORATORY – HRED
HQ USASOC
RDRL HRM CN R SPENCER
BLDG E2929 DESERT STORM DRIVE
FORT BRAGG NC 28310

1 (PDF) ARMY G1
DAPE MR B KNAPP
300 ARMY PENTAGON RM 2C489
WASHINGTON DC 20310-0300

<u>ABERDEEN PROVING GROUND</u>

12 (PDF) DIR USARL
RDRL HR
L ALLENDER
P FRANASZCZUK
K MCDOWELL
RDRL HRM
P SAVAGE-KNEPSHIELD
RDRL HRM AL
C PAULILLO
RDRL HRM AR
M BOYCE
RDRL HRM B
J GRYNOVICKI
RDRL HRM C
L GARRETT
RDRL HRS
J LOCKETT
RDRL HRS B
M LAFIANDRA
RDRL HRS D
A SCHARINE
RDRL HRS E
D HEADLEY

INTENTIONALLY LEFT BLANK.